

## **Data Flow Framework: A persona-based repository to modeling recommender systems**

*Rodrigo Dantas da Silva<sup>1</sup>, José Arilton Pereira Filho<sup>1</sup>, Philippi Sedir Grilo de Moraes<sup>1</sup>, Ricardo Alexsandro de Medeiros Valentim<sup>1</sup>, Karilany Dantas Coutinho<sup>1</sup>, Azim Roussanaly<sup>2</sup>, Anne Boyer<sup>2</sup> and Carlos Alberto Oliveira<sup>1</sup>*

Laboratory of Technological Innovation in Health (LAIS)<sup>1</sup>

Lorraine Research Laboratory in Computer Science and its Applications (LORIA)<sup>2</sup>

### **Track: Academic Research**

#### **1 Purpose**

The present project comes to propose a framework of data acquisition and structured repository in order to be able to apply machine learning techniques to content recommendation. For this, it is intended to use as a basis the modeling and prediction of future scenarios created from the AvaSUS (Virtual Learning Environment of Single Health System in Brazil) data, coverage of primary care in Brazil data, quality indicators from primary care, health facilities and health professional data, as well as data from diseases with mandatory notification. The AvaSUS data are obtained directly from the its database, whereas the others are part of a data lake built for research. The construction of a repository is due to the need of conformation of the structures, to this was chosen to work with the Experience API (or xApi) which defines an ontology structuring users and resources from one action.

#### **2 Design**

Connect space analysis and health data in order to provide a better understanding of the health-disease-care scenarios is not a recent issue. In the late 18th century, Leonhard Ludwig Finke published the book “*Versuch einer allgemeinen medicinisch-praktischen Geographie*”, which brought the worldwide diseases topography at that time. In the following century, John Snow contributed in fight against the cholera, mapping the water supply and infections. Hence, arise the social maps showing the correlation between poverty and cholera. Charles Booth’s works, in the end of that century, showed space correlations between social factors and distribution of health conditions. The relation between the geography of a locality, social factors and population health are recognized study field known as epidemiology.

The epidemiology studies application, as a consequence, yields a large amount of data sets and with this data there is a lot of possibilities. The data processing and correlation from multiple sources that assist health particular areas allow us to develop a perception from health geography in a locale. It brings us the possibility to realize how the health care services behaves and react to the local cases of diseases and global cases. Therefore, enable us to understand how and what happened. Nowadays, this approach is widely used at the epidemiological surveillance and healthcare strategic policies.

One of the possibilities noticed for those data is the healthcare professional (HP) lifelong learning. The usage of empirical evidences to this qualification model is already widespread around the world. The knowledge stemming from the scientific publications has been used to support the healthcare management, considering safety on interventions and the ethic in the totality of actions. Nevertheless, the scientific production of relevant contents in some health fields could take a long time, as scientific proposals in healthcare often require laboratory evaluations and human testing.

We propose a data flow framework extending the data lake architecture and the xApi, developing a data repository with multiple sources, considering data that goes beyond Learning Management System (LMS). It gets data that represents the reality of HP routine with the ability to represent existing requirement groups and then allowing the content recommendation from the real needs. The proposed framework also enables future works to use numerous machine learning models for different proposals, including real time learning. Thus, we are aiming that the proposed framework allows the Brazilian HP lifelong learning to be oriented by predicted future problems, making them previously qualified and prepared to the needs.

### **3 Results**

The proposal is to combine data from AvaSUS with data from the official systems of the Ministry of Health. By structuring this data in a standardized way in a repository, it is possible to create a link between the context and the characteristics of the users, thus allowing the modeling of predictive systems and content recommendation systems.

### **4 Implications**

Applying the latest techniques of data analysis can help the process of professional lifelong learning. Data science techniques, as data mining and machine learning, could help to predict the future scenarios, enable us to understand where the gaps from the current process are and then mapping the qualification requirements to overtake the predictions. The AvaSUS has started in 2015 with the aim of eliminating gaps in lifelong learning of the HP. It is the official

platform in Brazil maintained by Ministry of Health and used all over the country by HP and students to qualification. At this point, there are more than 380 thousand enrolments and 180 courses available in the most diverse areas of healthcare.

## **5 Acknowledgments**

This work was supported by the Laboratory of Technological Innovation in Health (LAIS) and Lorraine Research Laboratory in Computer Science and its Applications (LORIA).

## **6 References**

Anne Boyer, Gérard Casanova, Azim Roussanaly, Florence Ducreau. Pericles or How to Assess and to Improve Quality in Higher Education. EDEN Annual Conference 2015, Expanding Learning Scenarios, Jun 2015, Barcelone, Spain. Expanding Learning Scenarios, Opening Out the Educational Landscape. {hal-01256003}

Drachsler H., Verbert K., Santos O.C., Manouselis N. (2015) Panorama of Recommender Systems to Support Learning. In: Ricci F., Rokach L., Shapira B. (eds) Recommender Systems Handbook. Springer, Boston, MA

Lim, Kin Chew. "Case Studies of xAPI Applications to E-Learning." The Twelfth International Conference on eLearning for Knowledge-Based Society. 2015.

Ferguson, Rebecca (2012). Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6) pp. 304–317.